SAND2019-10291PE

Time-stepping Challenges, Opportunities, and Potential Solutions in the DOE's Energy Exascale Earth System Model (E3SM)

K. Chad Sockwell

Sandia National Laboratories is a multi-mission laboratory managed and operated by National Technology and Engineering Solutions of Sandia, LLC., a wholly owned subsidiary of Honeywell International, Inc., for the U.S. Department of Energys National Nuclear Security Administration under contract DE-NA-0003525.

August 29, 2019

Energy Exascale Earth System Model (E3SM)

Motivation

- Model built to perform on future exascale machines
- Finer resolution and multiresolution meshes
  - regions of interest
  - coastal refinement
  - coupling to tidal estuary models
- Model for Prediction Across Scales - Ocean (MPAS-O)
- High-Order Methods Modeling Environment (HOMME) -Atmosphere
- Coupling Approaches Next Generation Architectures (CANGA)

Challenges

- Time-stepping becomes bottleneck as resolution becomes finer
  - Explicit Methods - CFL
  - Implicit Methods - Global Communication
  - Efficiency per time-step is key $\rightarrow$ Time-to-solution
- Error is dominated by spatial error
- Long term intergation: stability, conservation properties, statistics
- Physcial Fluxes between models

Outline

Time-Stepping Solutions in MPAS-O

Exponential Time-Differicing

Time Stepping in HOMME

Model Coupling

High Resolution Model Spin-up

Multi-layer shallow water model

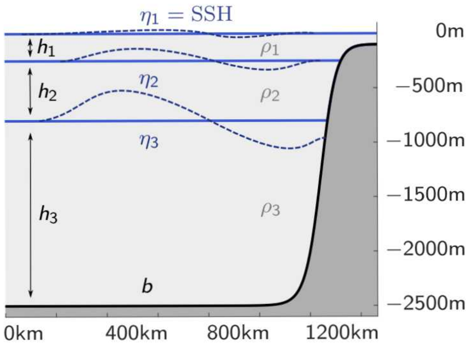
Model and solution variables

- L layers with uniform densities ρ_l, $l = 1, \ldots, L$
- Bottom topography b
- Layer thickness

$$h_l = \bar{h}_l + \Delta h_l$$

- Layer velocity u_l (along isopycnal)
- Layer interfaces

$$\eta_l = b + \sum_{i=l}^{L} h_i \ , \ \eta_{L+1} = b;$$



Mutli-layer Rotating Shallow Water Equations (MLRSWE)

Typically models use 60 to 100 layers
Layers coupled through gravity and pressure terms

$$\frac{\partial h_l}{\partial t} = -\nabla \cdot (h_l \vec{u}_l)$$

$$\frac{\partial \vec{u}_l}{\partial t} = -qh(\hat{l} \times \vec{u}_l) - \nabla \left(g[b + \sum_{j=1}^{l} \frac{\rho_j}{\rho_l} h_j + \sum_{j=l+1}^{L} h_j] \right) - \nabla K_l + \mathcal{F}_l \; ,$$

$$\vec{u}_l \cdot \vec{n} = 0 \text{ on } \Gamma, \; \forall \; l = 1, 2, \ldots, L$$

- Kinetic energy: $K_l = |\vec{u}_l|^2/2$
- Potential vorticity: $q_l(h_l, \vec{u}_l) = (\hat{k} \cdot \nabla \times \vec{u}_l + f)/h$
- Forcing: $\mathcal{F}(h, \vec{u})$ - wind, drag, diffusion,...
- Gravitational acceleration g, coriolis force parameter f, bottom topography b, unit vector in z direction $\hat{k}$
- Mimetic TRiSK scheme is used in space discretization

Multiple Time-Scales: Barotropic vs. Baroclinic

- Consider red term in matrix form (couples layers through pressure term)

$$R\vec{h} = \begin{pmatrix} 1 & 1 & 1 & \dots & 1 \\ \rho_1/\rho_2 & 1 & 1 & \dots & 1 \\ \rho_1/\rho_3 & \rho_2/\rho_3 & 1 & \dots & 1 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho_1/\rho_l & \rho_2/\rho_l & \rho_3/\rho_l & \dots & 1 \end{pmatrix} \vec{h}$$

- For $\rho_l \approx \rho$, R has one large eigenvalue$\rightarrow$ gravity wave (barotropic mode)
- Densisty range in ocean $1025 \leq \rho \leq 1028$
- Linearize at resting state $\bar{V} = (\bar{h}, 0)^\top$. Eigenvalue problem in vertical for three layers

$$\begin{aligned} h_t &= -\nabla \cdot (\bar{h}\, u) \\ u_t &= -g \nabla R\, h + f\, \vec{k} \times u \end{aligned} \quad \rightsquigarrow \quad \begin{aligned} \lambda h_\lambda &= -\nabla \cdot (\bar{h}\, u_\lambda) \\ \lambda u_\lambda &= -g \nabla R\, h_\lambda + f\, \vec{k} \times u_\lambda \end{aligned}$$

Multiple Time-Scales: Barotropic vs. Baroclinic

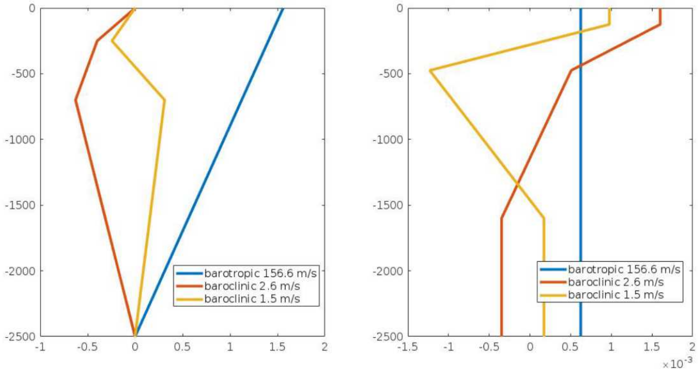


Figure: Vertical $\eta = h - b$ modes and velocity modes and corresponding wave speeds.

The Split-Explicit Method[1]

Fastest mode is the barotropic mode (variables $(\mathbf{h}, \mathbf{u})$)

$$u_{\mathrm{BT},l} \approx \mathbf{u} \,, \quad h_{\mathrm{BT},l} \approx \mathbf{h} \, \frac{h_l}{\sum_l h_l} = \mathbf{h} \, \frac{h_l}{H} \,, \quad l = 1, \ldots, L.$$

Define single-layer barotropic components

$$\mathbf{u} = \sum_l \frac{h_l}{H} u_l \,, \quad \mathbf{h} = \sum_l h_l$$

Barotropic Equations (Fast)

$$\frac{\partial \mathbf{h}}{\partial t} = -\boldsymbol{\nabla} \cdot (\mathbf{hu}) \,,$$
$$\frac{\partial \mathbf{u}}{\partial t} = -f \times \mathbf{u} - g \boldsymbol{\nabla} \mathbf{h} + G$$

Barotropic Equations (Slow)

$$\frac{\partial h_{\mathrm{BC},k}}{\partial t} = F_{\mathrm{RSWE},l} + \boldsymbol{\nabla} \cdot \mathbf{hu} \,,$$
$$\frac{\partial u_{\mathrm{BC},k}}{\partial t} = F_{\mathrm{RSWE},l} + f \times \mathbf{u} + g \boldsymbol{\nabla} \mathbf{h} + G$$

---

[1] **Higdon**; **2002**.

The Split-Explicit Method

Resuting Algorithm

- Explicity 3d solve: Large baroclinic time step Δt
- Explicity 3d solve: Subcycle barotropic Time Step $\Delta t/J$ from t^n to $t^{n+2} = t^n + 2\Delta t$, $J = 2m + 1$
- Average bartropic solution and define at $t^{n+1} = t^n + \Delta t$
- Combine barotropic and baroclinic components

Pros

- Time step increase by factor of ≈ 30 in MPAS-0
- Avoid costly global communications (layers a stacked on each processor)

Cons

- Small sub-steps require communication at every step (halo update)
- Must solve BT equations for $2\Delta t$ intervals
- Performance gain from solving 2d equations is reduced by latency

[inline]Ask Phil about time-step gain, what is BT and BC TS and how many subcylces? effecientcy gain?

The Split-Explicit Method: Possible Solutions

Use Less Explicit Time steps

Less BT time steps means less communication

Filtering

- Perform quadrature on BT time steps on smaller interval than $2\Delta t$
- Siddhartha Bishnu (LANL - FSU), Mark Peterson (LANL)
- Results in less barotropic time steps - Less communication

Local Time Stepping

- Multi-resolution time step are bound by smallest cells
- Thi-Thao-Phuong Hoang (AU), Wei Leng (CAS), Lili Ju (USC), Zhu Wang (US), Konstantin Pieper (ORNL)
- Results in less barotropic time steps - Less communication

The Split-Explicit Method: Possible Solutions

Take Larger BT time step

Possibly reduced amount of communication

Must perform global-reductions

Implict BT time step

- In recent years, cost of global reductions and lack of efficient preconditioners has turned MPAS-O away from fully implicit method
- Recent developments on preconditiors and reduced cost of glocal communication has sparked interest
- Phil Jones (LANL), Katherine Evans (ORNL), Postdoc? (ORNL)

Performance Gain

- Preconditioned Krylov methods
- Matrix-vector multiplications and dot products (global reductions), for large time step
- Many halo updates for explicit methods
- Can you beat explicit with good preconditioner for implicit method $\rightarrow$ Yes!

Performance Gain

What is accuracy cost of smoothing fast scale?

Another Solution: Exponential Runge-Kutta methods (Exponential Time-Differicing - ETD)[2]

Solve linear part "exactly" (stiffness is contained in linear part)

ETD Methods approximation gives control of energy dissipation

- split the forcing term in linear part and remainder

$$\begin{aligned}\partial_t V &= F(V)\\ &= AV + \big[F(V) - AV\big]\\ &= \underbrace{F(V_n) + A(V - V_n)}_{\text{Linear approximation}} + \underbrace{\big[F(V) - F(V_n) - A(V - V_n)\big]}_{\text{Residual } R(V)}\end{aligned}$$

- Variation of constans formula

$$V(t^{n+1}) = V(t^n) + \int_{t^n}^{t^{n+1}} \exp(\Delta t A)(F(V(t)) + A(V(t) - V(t^n)))\,\mathrm{d}t$$

- Exponential RK2 method (for $A \neq F'(V_n)$)

$$\begin{aligned}V_n^1 &= V_n + \Delta t \varphi_1(\Delta t A) F(V_n) && \text{“Exponential Euler”}\\ V_{n+1} &= V_n^1 + \Delta t \varphi_2(\Delta t A) R(V_n^1) && \text{“Second order correction”}\end{aligned}$$

- φ-functions:

$$\varphi_0(z) = \exp(z), \quad \varphi_1(z) = (\exp(z) - 1)/z, \quad \varphi_2(z) = (\exp(z) - 1 - z)/z^2, \quad \ldots$$

[2] **hochbruck2010exponential**.

Choice of the linear operator

Splitting of the forcing term at step n

$$F(V) \approx F(V_n) + A_n(V - V_n)$$

- Wave operator approximation:
  linearize at $\bar{V} = (\bar{h}, 0)$ for a reference heights $\bar{h}$, zero velocities

- Hamiltonian formulation

$$F(V) = J(V)\frac{\delta H}{\delta V}$$

$$F'(V) = J'(V)\frac{\delta \mathcal{H}}{\delta (V)} + J(V)\frac{\delta^2 \mathcal{H}}{\delta (V)^2}$$

$$F'(\bar{V}) = J'(\bar{V})\frac{\delta \mathcal{H}}{\delta V} + J(\bar{V})\frac{\delta^2 \mathcal{H}}{\delta (V)^2} = J(\bar{V})\frac{\delta^2 \mathcal{H}}{\delta V^2}$$

Multilayer rotating wave equation

$$\partial_t V = AV \qquad \rightsquigarrow \qquad \begin{cases} \partial_t h = -\boldsymbol{\nabla} \cdot \left(\bar{h}\, u\right) \\ \partial_t u = -g \boldsymbol{\nabla} R\, h + f\, k \times u \end{cases}$$

Reduction to barotropic mode (similar to split-explicit scheme[4])

- Fastest mode is the barotropic mode (variables $(\mathbf{h}, \mathbf{u})$)

$$u_{\mathrm{BT},k} \approx \mathbf{u}, \quad h_{\mathrm{BT},k} \approx \mathbf{h} \, \frac{\bar{h}_k}{\sum_k \bar{h}_k}, \quad k = 1, \ldots, L.$$

- corresponding Ansatz[3]

$$\begin{pmatrix} h_{\mathrm{BT},l} \\ u_{\mathrm{BT},l} \end{pmatrix} = G \begin{pmatrix} \mathbf{h} \\ \mathbf{u} \end{pmatrix}$$

- corresponding left inverse with $G^\dagger G = \mathrm{Id}$

$$\begin{pmatrix} \mathbf{h} \\ \mathbf{u} \end{pmatrix} = G^\dagger \begin{pmatrix} h \\ u \end{pmatrix}$$

- barotropic projection

$$P = GG^\dagger$$

- replace multilayer operator A by projected version

$$PAP = G \underbrace{G^\dagger A G}_{=\boldsymbol{A}} G^\dagger$$

[3] **Konstantin Pieper**; **K. Chad Sockwell**; **Max Guznburger**; **2019**.
[4] **Higdon**; **2002**.

Construction of the projection

- general Ansatz/parameterization:

$$\begin{pmatrix} \hat{h} \\ \hat{u} \end{pmatrix} = G \begin{pmatrix} \boldsymbol{h} \\ \boldsymbol{u} \end{pmatrix}$$

- for any $V = (h, u)$, find PV and $\boldsymbol{V} = G^{\dagger} V$ by

$$PV = GG^{\dagger} V = G\boldsymbol{V}$$
$$\text{where } \boldsymbol{V} = \arg\min_{\boldsymbol{V}} \underbrace{\| V - G\boldsymbol{V} \|_{M\delta^2 H}}_{\substack{\text{reconstruction error} \\ \text{in energy norm}}}$$

- $G^{\dagger}$ has a closed form solution that can be computed in each cell/edge stack (since $M \frac{\delta^2 H}{\delta V^2}$ diagonal)
- for G the barotropic Ansatz, similar $G^{\dagger}$ appears in split-explicit scheme

Barotropic method

- consider ETD-methods based on $A = J(\bar{V})\frac{\delta^2\mathcal{H}}{\delta V^2}$ with

$$\partial_t V = F(V) = PAP\,V + r(V)$$

- barotropic projection $P = GG^\dagger$
- Linear operator retains structure $\boldsymbol{A} = G^\dagger A G = \boldsymbol{J}\,\frac{\boldsymbol{\delta^2\mathcal{H}}}{\boldsymbol{\delta V^2}}$

$$\boldsymbol{\delta^2 H} = G^\top \delta^2 H(\bar{V}) G$$
$$\boldsymbol{J} = G^\dagger J(\bar{V})(G^\dagger)^\top$$

- can be tweaked to be either mass or volume conserving
- computation of the φ-functions

$$\varphi_s(\Delta t PAP) = \frac{1}{s!}(\mathrm{Id} - P) + G\varphi_s(\Delta t\boldsymbol{A})G^\dagger$$

- Barotropic exponential Euler (B-ETD)[5]

$$V_{n+1} = V_n + \Delta t(\mathrm{Id} - P)F(V_n) + \Delta t\, G\varphi_1(\Delta t\boldsymbol{A})G^\dagger F(V_n)$$

[5] **Konstantin Pieper**; **K. Chad Sockwell**; **Max Guznburger**; **2019**.

Krylov-methods for φ-functions of skew-symmetric operators

- fix A and b: build the optimal polynomial p_K with

$$\varphi_s(\Delta t A)b \approx p_K(\Delta t A)b$$

- Krylov space

$$\mathcal{K}_K(A, b) = \{ A^k b \mid k = 0, \dots, K \}$$

- Arnoldi-process: orthogonal basis V_K of $\mathcal{K}_K(A, b)$ (cost quadratic in K)

$$V_K^\top A V_K = H_K \in \mathbb{R}^{K \times K}$$

- dense evaluation of $\varphi_s(\Delta t H_K)$ (cost independent of N_{dof})

skew-symmetry of $\bar{A} = J(\bar{V}) \frac{\delta \mathcal{H}}{\delta V}$

$$M_H A = -A^\top M_{\mathcal{H}} \quad \text{for} \quad M_{\mathcal{H}} = M \delta^2 \mathcal{H}(\hat{V})$$

- skew-Lanczos process with respect to the M_H inner product (cost linear in K)
- H_K tri-diagonal and skew-symmetric

Number of Krylov vectors

- fix A and b: build the optimal polynomial p_K to approximate

$$\varphi_s(\Delta t A)b \approx p_K(\Delta t A)b$$

- interpolates $\varphi_s(z)$ on the eigenvalues of H_K (imaginary, due to skew-symmetry)

$$V_K^\top A V_K = H_K \in \mathbb{R}^{K \times K}$$

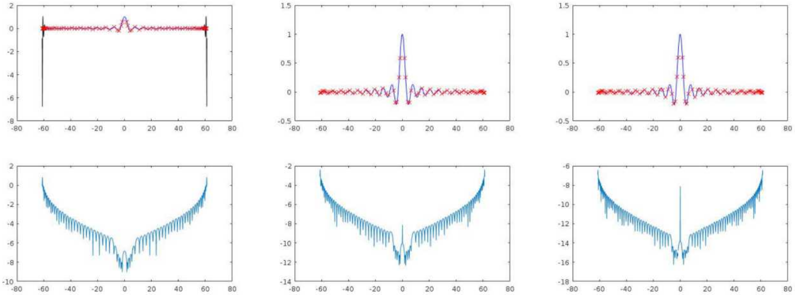


Figure: Real part of $p_K(z)$ for $K = 70$, $K = 80$, and $K = 90$
Corresponding minimum number of substeps for RK4: $4 \cdot N_{\mathrm{substep}} = 100$

Performance

3 Layer Sequential Test Case

With 3rd Order B-ETD method $\Delta t = 12.5\Delta t_{\mathrm{RK4}} = 34.5\Delta_{\mathrm{CFL}}$, 56 Krylov Vectors per φ_s evaluation (3 in total)

5 times speedup against standard RK4, expected to increrase with number of layers

Parallelization is required for fair comparison to split-explicit method

Hope that preservation of high frequencies is worth cost

Improvements

Perform Domain Decomposition and Krylov iteration is local

Combine local time stepping and Domain Decomposition

Could splitting be better? I would expect $60\Delta_{\mathrm{CFL}}$ if splitting was perfect

Could a more physical or geometrical approach give better splitting?

Time Stepping in HOMME

Hydrostatic Model

- ρ varies more than in the ocean. Wave-speed difference in factor of 3
- Vertical advection is stronger than in the ocean
- This leads to Horizontal-Explicit Vertical-Implicit methods (HEVI)- No communication required for solves $\rightarrow$ IMEX methods

Non-Hydrostatic Atmosphere

- Vertical Acoustic waves become fastest wave
- IMEX methods with HEVI-type splitting- No communication required for vertical solves[6]
- Exponential Time-Differing Methods for vertical vertical component (Andrew Steyer (SNL), Cassidy Krause (SNL)), (Sara Calandrini (FSU), Konstantin Pieper (ORNL), Max Gunzburger (FSU))

Semi-Lagrangian Tracer Transport[7]

- Mitigate CFL by solving remappnig problem
- Shape preseration, positivity, range perservation
- Minimize communication

---

[6] **Andrew Steyer (SNL)**; **et al. 2019**.

[7] **Andrew Bradley**; **et al. 2019**.

Model Coupling

Coupling Approaches for Next Generation Architectures (CANGA)

Currently Coupling is similar to one-step alternating Schwarz.

Time InteGration for Greater E3SM Robustness (TIGGER)

Pavel Bochev (SNL), Kara Peterson (SNL), Paul Kuberry (SNL), Nat Trask (SNL), K. Chad Sockwell (SNL)

How to non-intrusiviely coupling model components

Enforce BC's with Lagrange multipliers

Desires

Open question: Unconditionally stable, conservative coupling scheme

Ideally works with any choice of time-integrators across compoents, and for differently sized time-steps

Question

Could vartiational viewpoint give more insight into coupling, even if methods must be intrusive? Conservation and (conditionally) stability are a must

Model Spin-up

Ocean Equilibration

Developing high resolution initial condition for ocean - Grand Challenge

Bad initial condition leads to transients which are not from the forcing

Deep ocean currents have slow response to forcing

Roughly 1000 year spin-up to equilbrate the ocean

Possible Solutions

Developing Structure-Preserving reduced order model (K. Chad Sockwell (SNL), Luke Van Roekel (LANL), Andy Salinger (SNL) ,Konstatin Pieper (ORNL), Max Gunzburger (FSU))

Find solution space where equalibrated solution lives

Pose as optimal control problem

Could more physically inspired methods lead to solution?

Questions

Where do variational discretizations fit into climate models?

Are they efficient enough?

What advantages do they bring?

ROM for variational discretizations instead of ROM for mimetic discretizations?